Дистанційне навчання

До уваги студентів спеціальності «Транспортні технології (автомобільний транспорт)», I курс (скорочений термін навчання), група 1906.

У зв'язку з дистанційною формою складання літньої сесії 2019-2020 р.р. з 18 травня 2020 р. (наказ від 7.04 2020р. на сайті НУБіП) внесено такі коригування щодо дисципліни «Теорія ймовірності і математична статистика» (щоб встигнути за графіком деканатів, яких ще немає)

1. Рейтингова система складатиметься з 2 модулів (тест за модулем 2 складаєте за графіком).
2. Буде ще два завдання 7 та 8, матеріал яких буде входити до підсумкового контролю (заліку). Видача цих завдань буде на наступному тижні.
3. Прошу надсилати виконані завдання за адресою: wnyrk15@gmail.com

З цих завдань складатиметься оцінка за навчальну роботу.

4. Студентам, які не пройшли тестування за модулем 1, за модулем 2 та з підсумкового контролю (залік)– буде виставлено оцінку «незараховано».

5. Дату заліку назначаю на 12 травня 2020 р. Залік будемо проводити на ЕНК ВМ Т. Підсумковий тест з дисципліни "Вища математика", 4 семестр (30 тестів по 1 балу на протязі 1год. 20 хв.) Відомість маю закрити наступного дня (див. положення на сайті НУБіП).

Конкретне повідомлення буде пізніше.

Доц. Панталієнко Л.А.

Графік навчальної роботи за модулем 2 на ЕНК
(Завдання на сайті кафедри вищої та прикладної математики)

Група	Дисципліна	завдання №4 (дата здачі)	завдання №5 (дата здачі)	завдання №6 (дата здачі)	Тест за модулем 2 (дата здачі)
ТрТ1906.	Теорія ймовірності і математична статистика	до 3квітня 2020	до 10квітня 2020	до 20 квітня 2020	23 квітня 2020

Викладач: доц. Панталієнко Л.А.

Термін виконання Завдання №7: до 1 травня 2020р.

Прошу надсилати виконані завдання за адресою: wnyrk15@gmail.com

Завдання №7. За даними спостережень неперервної ознаки *X* генеральної сукупності необхідно: а) скласти статистичний розподіл вибірки, розбивши значення *X* на 5 інтервалів; б) побудувати гістограму відносних частот.

№ спостереження	Варіант									
	1	2	3	4	5	6	7	8	9	10
	X	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
1	0,2	12	4	11	1,2	0,2	7,1	1,1	12	1,2
2	3,8	15	5,1	13,2	1,5	1,1	7,4	2	11	2
3	4,1	14	4,2	17	2	3	7,9	2,2	8	5
4	4	12	2,7	16,1	3	4,2	8,1	2,8	7,3	4
5	4,2	11	3	15	4	2,1	7,2	3	2	3
6	5	12	3,8	19	4,2	4,8	7,5	1,2	3	6
7	5,8	15	4,9	20,2	5	4,2	4,9	9	3,8	7,5
8	6,2	13,2	6,1	12	3	4,4	3,2	9,3	3,5	8
9	3,8	11	1,1	13,1	3,2	6,1	8,1	2,8	6	8,3
10	3,3	10,8	1,7	12,2	4	6	4	2,7	6,2	8
11	4,7	9,2	2,8	11,1	8,1	5	5,3	2	5,3	3,8
12	2	9	7	12,5	8	4,7	4,5	3	5	2,7
13	1	9,1	5,8	12,6	6,5	8,2	5	5	7	5,1
14	0	8,7	3,5	9	6,1	3	5,3	6	7	5
15	2	8,2	4,2	9,5	4,9	4,1	5,5	6,2	7,9	8
16	2,7	8	8,1	9,8	10,2	2	11	7	7,2	9
17	3	5	6,2	7,8	5,3	6,1	11,2	7,2	8	7
18	3,2	5,8	6,3	2,2	8	5	7	8	8,1	10
19	2,5	9,2	5,1	7,1	7	5,2	9	9	9	2
20	8	2	5	8	6,2	4,9	9,5	1,2	5,8	3

№ спостереження	Варіант									
	11	12	13	14	15	16	17	18	19	20
	X	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
1	1,1	0,2	4	8	18,5	1,1	17	2,3	10	13
2	2,3	0,3	1	9	19	2	12	4,2	11	14,1
3	5,1	1,4	0	10	18	3,2	16,2	5,1	13,2	11,5
4	2,6	1	3	11	17,3	4,1	11,1	5,8	9	12,1
5	2,9	2	2	10,6	16,8	5	14	5,2	8,2	17,2
6	3,1	2,2	1,2	13,2	15,1	5,8	14,2	5,8	7,8	10,9
7	4	1,6	3,1	11	20,2	9,1	17,1	3,5	9,1	12,1
8	4,2	1,9	0,8	11,1	20,5	8,2	17,3	5,1	6,9	12,3

9	3,8	1,8	3	10,8	19,2	9,1	19,1	7	7,1	14
10	3,7	2,4	3,1	9	18	9,3	20,2	6,8	7,3	11
11	3,3	1,8	2,9	8,5	17,3	10	21,3	4,9	6,5	10
12	2,1	0,9	2,8	6,2	16,8	8,2	23,1	6,5	6	12
13	2,8	2	1,9	9	15	7,7	23	6,2	5,8	11
14	2,9	2,1	1,8	9,2	14,5	6,2	22	8,3	5,2	9,2
15	3	1,5	3,9	9,8	19,1	3	20	5	4	8,7
16	3,2	1,7	3,2	10,1	18	7,5	20,1	5,8	4,5	9,8
17	3,5	2,5	3	10	18,3	9,2	19	4,8	8	5,2
18	4	3,2	2,8	11,5	15,8	12	19,2	4,1	8,2	12
19	4,4	3,1	2	10,1	16,5	12	16	4	7	13,2
20	4,7	2,5	2,5	9	17	16,1	14	2,8	2	17

№ спостереження	Варіант									
	21	22	23	24	25	26	27	28	29	30
	X	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
1	12,1	9,1	10,2	10	5,9	9	6,1	7,3	8,5	8,9
2	8	7,8	7,5	6,5	5,3	5,5	5,6	4,5	6,4	7
3	11	8,8	9,5	9	5,8	8,1	2,2	7	8,4	8,7
4	9,4	7,1	10	3,5	2,2	5	7	6,1	7	5,4
5	5,2	8,1	7,8	8,5	8	7	3	5	4,8	10,2
6	12	9	10,4	6	6	5,1	8,4	5,8	9,6	8,6
7	6	8	6,2	4	4	3,1	5,2	7,5	8	7,5
8	17	8,4	9	7,2	6,2	8,2	4,9	5,5	8,1	9
9	8,2	7,7	8	8	4,2	6,7	7,2	7,8	7,2	8
10	7,5	8,3	11	7	5,7	6,6	6	6,5	9	9,5
11	14,8	8,6	7	5	3	8,5	4,3	5,2	8,2	10
12	11,5	7,5	9,2	7,8	6,9	7,9	7,4	8	7,6	8,5
13	9,8	8,9	8,8	8,1	6,8	6	2,8	6,6	10	11,8
14	15	8,7	11,5	5,5	5,1	10	7	6,3	8,7	10,1
15	11,8	8,2	10,1	10,5	7,8	7,8	5	6,7	10,2	7,8
16	10	10	7,9	8,3	5	6,5	7,8	6	7,9	11
17	9	8,5	11,1	6,8	8,5	10,1	8,8	8,1	11	10,5
18	16	9,4	8,9	12	5,4	8	8	6,8	11,1	10,8
19	12,3	9,8	13,2	9,2	10,2	10,2	11,2	8,4	12,8	12
20	20,3	10,1	12	12,5	7,4	12	11	9,5	12	13,4

Теоретичний матеріал до завдання №7 і до іспиту.

6.1. Первинна обробка статистичних даних.

В результаті спостережень та реєстрації масових випадкових явищ отримуються статистичні дані або статистичний матеріал (це, наприклад, похибки різних вимірювань). Встановлення закономірностей, яким підпорядковані масові випадкові явища, ґрунтується на вивченні методами теорії ймовірностей статистичних даних.

Нехай потрібно дослідити сукупність однорідних об’єктів щодо деякої якісної або ж кількісної ознаки X, що характеризує ці об’єкти. Вважаємо, що ознака, яка вивчається є випадковою величиною

Генеральною сукупністю називають всю сукупність об’єктів, що підлягають вивченню. Сукупність n об’єктів, випадково відібраних з генеральної сукупності, називається вибірковою сукупністю або вибіркою. Число n − об’єм вибірки.

Одним із основних методів статистичного дослідження є вибірковий метод, суть якого полягає в тому, що на основі вивчення вибірки, що складається зі скінченого числа n об’єктів, робляться певні висновки для всієї генеральної сукупності. Природно, що ці висновки та оцінки відображують випадковий характер зібраних статистичних даних і тому повинні вважатись наближеними оцінками ймовірнісного характеру.

Нехай для вивчення кількісної ознаки X з генеральної сукупності вибрано вибірку об’єму n, причому при n незалежних випробуваннях ознака X прийняла значення x_1 – n_1 разів, x_2 – n_2 разів, ..., значення x_k – n_k разів; $n_1 + n_2 + ... + n_k = n$.

Значення x_i, $i = 1,2,...,k$ ознаки X називають варіантами, а послідовність варіант у зростаючому порядку – варіаційним рядом.

Числа n_i (кількість спостережень) називаються частотами, а числа $\omega_i = \dfrac{n_i}{n}$ – відносними частотами варіант x_i ($i = 1,2,...,k$); причому $\sum_{i=1}^{k} \omega_i = 1$.

Статистичним розподілом вибірки називається перелік варіант і відповідних їм частот або відносних частот. Статистичний розподіл можна задавати також у вигляді послідовності інтервалів зміни варіант і відповідних цим інтервалам частот або відносних частот (в якості частоти певного інтервалу приймають суму частот варіант, що потрапили в цей інтервал). В результаті обробки статистичних даних одержується дискретний або інтервальний варіаційний ряд.

Дискретний варіаційний ряд випадкової величини X подається у вигляді таблиці:

Таблиця 1.

X	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
ω_i	ω_1	ω_2	…	ω_k

Приклад 6.1. Заданий розподіл частот ДВВ X об'єму $n = 20$:

X	2	6	12
n_i	3	10	7

Скласти розподіл відносних частот.

Розв'язання. $n = 3 + 10 + 7 = 20$. Знайдемо відносні частоти:

$$\omega_1 = \frac{n_1}{n} = \frac{3}{20} = 0{,}15;\ \omega_2 = \frac{n_2}{n} = \frac{10}{20} = 0{,}5;\ \omega_3 = \frac{n_3}{n} = \frac{7}{20} = 0{,}35.$$

Контроль: $0{,}15 + 0{,}5 + 0{,}35 = 1$. Результати обчислень зводимо в таблицю:

X	2	6	12
ω_i	0,15	0,5	0,35

Якщо вивчається неперервна випадкова величина X і число спостережень відносно велике ($n \geq 20$), будується інтервальний варіаційний ряд. Для цього проміжок, в якому містяться всі результати спостережень вибірки, розбивають на кілька (від 5 до 15) рівних або нерівних інтервалів, підраховують суми $n_1, n_2, ... n_r$ частот варіант, що потрапили в ці інтервали, відносні частоти $\omega_i = \frac{n_i}{n}$ та щільності відносних частот $p_i^* = \frac{\omega_i}{|\Delta_i|}$ ($|\Delta_i|$ – довжина i-го інтервалу, $i = 1, 2, ..., r$).

Довжина кожного інтервалу, як правило, береться сталою величиною $h = \frac{x_{\max} - x_{\min}}{M - 1}$, $M = [1 + 3{,}32 \lg n]$ (n – об'єм вибірки). Межі послідовних інтервалів позначають $x_0, x_1, ... x_M$, кількість інтервалів – M, а самі інтервали

$$[x_0, x_1[,\ [x_1, x_2[,\ [x_2, x_3[,\ ...,\ [x_{M-1}, x_M[, \text{ де } x_0 < x_1 < x_2 < ... < x_M.$$

Звичайно, значення, що знаходяться на межі інтервалів, відносять до правої межі інтервалу.

Інтервальний варіаційний ряд випадкової величини X подається у вигляді таблиці:

Таблиця 2.

Інтервали	$[x_0, x_1[$	$[x_1, x_2[$	…	$[x_{M-1}, x_M[$
n_i	n_1	n_2	…	n_r
ω_i	ω_1	ω_2	…	ω_r
p_i^*	p_1^*	p_2^*	…	p_r^*

Межі часткових інтервалів визначають за формулами:

$$x_0 = x_{\min} - \frac{h}{2},\; x_1 = x_0 + h,\; x_2 = x_1 + h, \ldots, x_i = x_{i-1} + h, \ldots, x_M = x_{M-1} + h = x_{\max} + \frac{h}{2}.$$

Отже, розбиття на інтервали здійснюється таким чином, щоб найменше значення вибірки потрапляло у середину першого інтервалу, а найбільше значення вибірки – у середину останнього.

Зауваження. Для неперервної ознаки X закон розподілу можна будувати також у вигляді таблиці 1, вибираючи в якості x_i, $i = 1,2,...,k$ середнє значення з відповідного часткового інтервалу.

Приклад 6.2. В результаті 20-ти вимірювань деякої фізичної величини одним приладом (без систематичних похибок) одержано такі результати: 18,50; 15,97; 18,81; 18,93; 17,16; 19,04; 21,58; 18,50; 14,35; 13,21; 14,96; 19,25; 12,12; 14,55; 14,30; 14,39; 14,24; 18,48; 16,38; 16,45.

Скласти статистичний розподіл вибірки, розбивши увесь діапазон на $k = [\,1 + 3{,}32 \lg n\,]$ інтервалів (n − об'єм вибірки).

Розв'язування. 1) Запишемо послідовність варіант у зростаючому порядку, тобто варіаційний ряд
12,12; 13,21; 14,24; 14,30; 14,35; 14,39; 14,55; 14,96; 15,97; 16,38; 16,45; 17,16; 18,48; 18,50; 18,50; 18,81;18,93; 19,04; 9,25; 21,58.

Як бачимо, діапазон зміни варіант у вибірці становить 12-22. Далі увесь діапазон розбивається на декілька (часткових) інтервалів, кількість яких обчислюється за формулою $k = [1 + 3{,}32 \lg n\,]$. У нашому випадку $n = 20$, а $k = [5{,}316] = 5$. Отже, інтервалів буде п'ять: 12-14, 14-16, 16-18, 18-20, 20-22 однакової довжини $h = \frac{22-12}{5} = 2$.

Визначимо частоти варіант вибірки для кожного часткового інтервалу. У перший інтервал 12-14 потрапляє 2 значення (12,12; 13,21). Тому $n_1 = 2$; у другий ,14-16, − 7 значень (14,24; 14,30; 14,35; 14,39; 14,55; 14,96; 15,97), тому $n_2 = 7$. Аналогічно: $n_3 = 3, n_4 = 7, n_5 = 1$.

Подамо інтервальний варіаційний ряд випадкової величини у вигляді таблиці :

Інтервали	[12-14[	[14-16[	[16-18[	[18-20[	[20-22[
n_i	2	7	3	7	1

Контроль: 2+7+3+7+1=20 (сума всіх частот дорівнює об'єму вибірки).

6.2. Зображення результатів групування.

Для наочності статистичний розподіл вибірки зображують у вигляді графіків чи, там де доцільно, діаграм.

Діаграмою називають рисунок, на якому результати для дискретних чи неперервних даних представлені однотипними геометричними об’єктами, а їхні частоти чи відносні частоти – геометричними характерами цих об’єктів.

Розрізняють три основні види діаграм: стовпчикову, секторну та лінійчату.

Стовпчикова діаграма – це рисунок, на якому результати із ряду розподілу чи інтервали групування представлені у горизонтальних чи вертикальних стовпчиках, відкладених від одного рівня, а довжини цих стовпчиків пропорційні частотам або відносним частотам цих результатів. Коефіцієнт пропорційності визначається за вибором масштабної одиниці.

Приклад 6.3. Для рядів розподілу, що відповідають розподілу за статтю та темпераментом (дискретні дані):

Д (дівчинка)	Х (хлопчик)
0,49	0,51

Ф (флегматик)	М (меланхолік)	Х (холерик)	С (сангвінік)
0,48	0,12	0,2	0,2

стовпчикові діаграми представлені на рис. 6.1, 6.2.

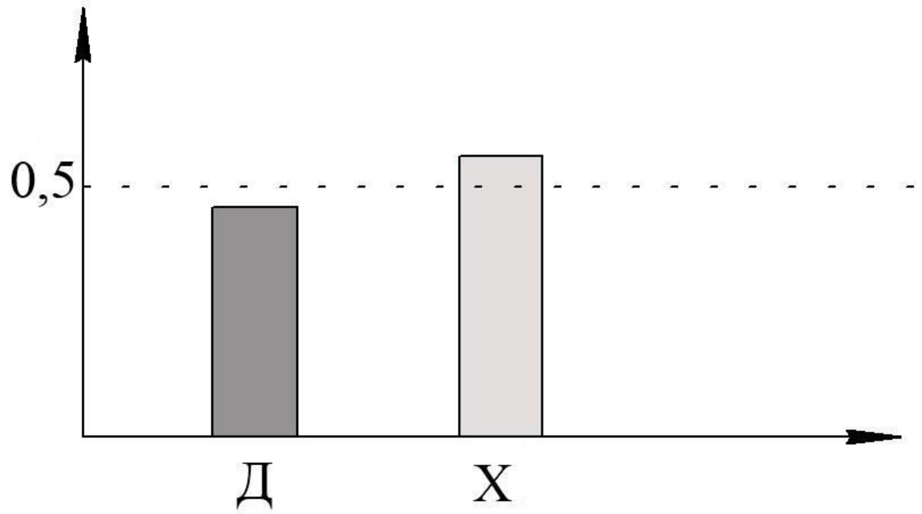

Рис. 6.1.

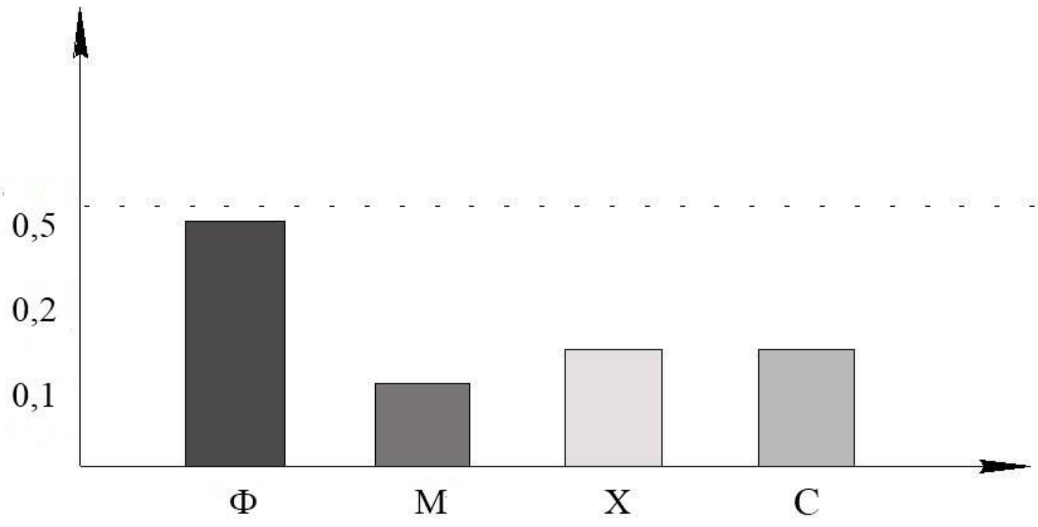

Рис.6.2.

Групову вибірку, що представлена таблицею,

Інтер вали	[50-80[	[80-110[	[110-140[	[140-170[
ω_i	0,1	0,3	0,5	0,1

можна зобразити у вигляді стовпчикової діаграми (рис. 6.3).

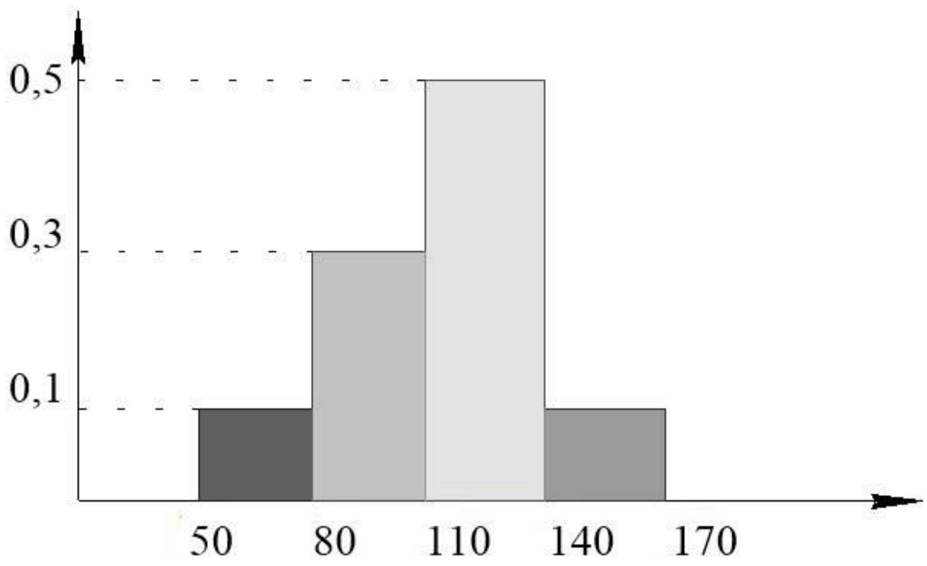

Рис. 6.3.

Секторна діаграма – це зображення ряду розподілу чи групованої вибірки, на якому результати (дискретні дані) чи групи результатів (неперервні дані – групована вибірка) представлені секторами одного кола чи циліндра, а кути цих секторів пропорційні частотам або відносним частотам цих результатів (рис.6.4).

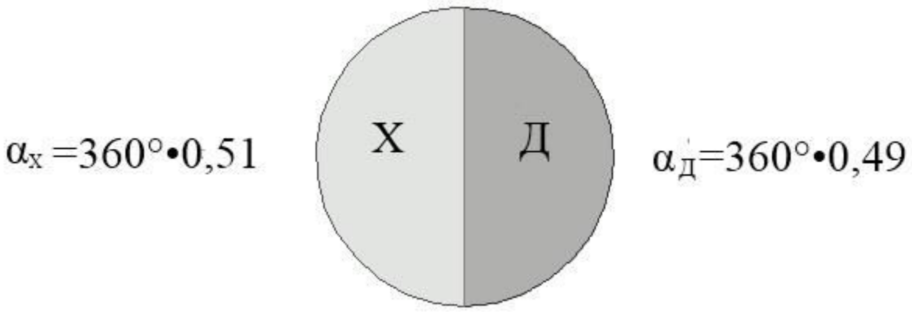

Рис. 6.4.

Лінійчата діаграма – це зображення ряду розподілу чи групованої вибірки у вигляді послідовних частин одного і того самого відрізка довжиною L.

Довжини відрізків l_i, що зображують результати чи групи результатів, пропорційні частотам або відносним частотам цих результатів: $l_i = L \cdot \omega_i$. Так, для рядів розподілу та групованої вибірки (приклад 6.3) відповідні діаграми мають вигляд:

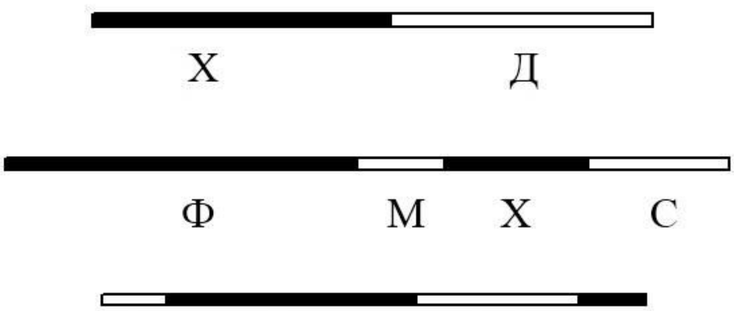

Рис. 6.5.

Графічним представленням для ряду розподілу у випадку дискретних числових даних є полігон частот, а у випадку неперервних – гістограма.

Полігоном частот називають ламану, відрізки якої з’єднують точки (x_1,n_1), (x_2,n_2), , (x_k,n_k). При цьому в прямокутній системі координат по осі абсцис слід відкласти значення x_1, x_2, ..., x_k варіант, а по осі ОУ – частот $n_1,n_2,...,n_k$. Аналогічно будується й полігон відносних частот.

Так, для ряду розподілу, що має числові результати

X	3	4	7	8	9
ω_i	0,1	0,2	0,3	0,1	0,3

полігон частот зображений на рис.6.6.

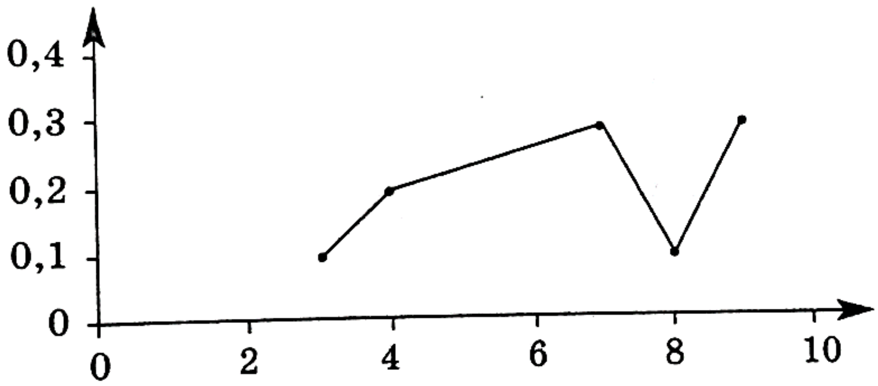

Рис. 6.6.

Функцію, визначену на всій числовій прямій, яка є кусково-сталою зі значенням $\frac{n_i}{h}$ ($\frac{\omega_i}{h}$) на інтервалах групування та нульовою – поза ними, називають гістограмою частот (відносних частот).

Для побудови гістограми по осі OX відкладають частинні інтервали, а над ними проводять відрізки, паралельні осі абсцис на відстані $p_i^* = \frac{n_i}{h}$ (або $\frac{\omega_i}{h}$), $i = 1,2,...,k$ відповідно (рис. 6.7). Площа i - того частинного прямокутника на гістограмі частот дорівнює $hn_i / h = n_i$ − сумі частот варіант i-го інтервалу, а площа гістограми частот дорівнює n − об'єму вибірки (площа гістограми відносних частот дорівнює одиниці).

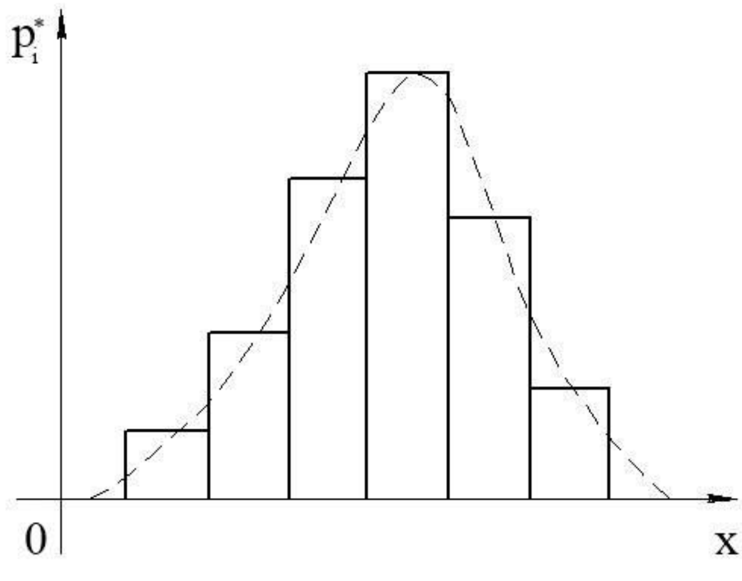

Рис.6.7.

Графічно гістограму зображують у вигляді графіка, що є сходинковою лінією, до якої належать верхні границі стовпчикової діаграми, коли стовпчики розташовані над відповідними інтервалами групування, а їх висотами є $\frac{n_i}{h}$ ($\frac{\omega_i}{h}$).

Якщо довжини частинних інтервалів малі, а об'єм вибірки n великий, то гістограма відносних частот слугує наближенням для щільності випадкової величини X (крива на рис. близька до кривої розподілу величини X).

Для побудови секторної діаграми зробимо додаткові розрахунки для центральних кутів за формулою $\alpha_i = 360^\circ \cdot \omega_i$ або $\alpha_i = 360^\circ \cdot \frac{n_i}{n_i}$.

Таким чином, маємо $\alpha_1 = 36^\circ = \alpha_3 = \alpha_5 = \alpha_6 = \alpha_7$, $\alpha_2 = 360^\circ \cdot 0{,}2 = 72^\circ$,

$\alpha_Ф = 360^\circ \cdot 0{,}48 = 172{,}8^\circ$; $\alpha_М = 360^\circ \cdot 0{,}12 = 43{,}2^\circ$; $\alpha_Х = 360^\circ \cdot 0{,}2 = 72^\circ$; $\alpha_С = 360^\circ \cdot 0{,}2 = 72^\circ$.
За цими даними одержимо секторну діаграму такого вигляду

Для побудови секторної діаграми зробимо додаткові розрахунки для центральних кутів за формулою $\alpha_i = 360^\circ \cdot \omega_i$ або $\alpha_i = 360^\circ \cdot \frac{n_i}{n_i}$.

Таким чином, маємо: $\alpha_1 = 36^\circ = \alpha_3 = \alpha_5 = \alpha_6 = \alpha_7$, $\alpha_2 = 360^\circ \cdot 0{,}2 = 72^\circ$, $\alpha_4 = 360^\circ \cdot 0{,}3 = 108^\circ$. Контроль: сума всіх кутів дорівнює 360°.

За цими даними одержимо секторну діаграму вигляду

6.3. Емпірична функція розподілу.

Нехай вивчається кількісна ознака X з невідомою функцією розподілу $F(x)$. За даними вибірки x_1, x_2, .., x_n об'єму n побудований статистичний розподіл випадкової величини X.

Для будь-якого дійсного числа x позначимо через n_x число варіант, менших x, тобто частоту події $X < x$.

Означення. Емпіричною функцією розподілу (функцією розподілу вибірки) називається функція

$$F^*(x) = \frac{n_x}{n},$$

яка визначає для кожного значення x відносну частоту події $X < x$.

На відміну від емпіричної функції розподілу вибірки функцію розподілу $F(x)$ генеральної сукупності називають теоретичною функцію розподілу. Різниця між емпіричною і теоретичною функцію розподілу полягає в тому, що теоретична функція $F(x)$ визначає ймовірність появи події $X < x$, а емпірична функція − відносну частоту цієї події.

Емпірична функція $F^*(x)$ вибірки слугує для оцінки теоретичної функції розподілу $F(x)$ генеральної сукупності величини *X*.

Властивості емпіричної функції розподілу.

1. Значення емпіричної функції розподілу $F^*(x)$ належать відрізку $[0;1]$:

$$0 \le F^*(x) \le 1.$$

2. Емпірична функція розподілу є неспадною функцією:

$$\text{при } x_1 < x_2 \;\; F^*(x_1) \le F^*(x_2).$$

3. Якщо x_1 − найменша варіанта, то $F^*(x)=0$ при $x \le x_1$;
   якщо x_k − найбільша варіанта, то $F^*(x)=1$ при $x > x_k$.

Приклад 6.4. За даним розподілом вибірки

X	1	4	6
n_i	10	15	25

побудувати емпіричну функцію розподілу $F^*(x)$.

Розв'язання. Знайдемо об'єм вибірки: $n = 10 + 15 + 25 = 50$. Найменша варіанта дорівнює 1, а найбільша − 6: $x_1 = 1$, $x_3 = 6$. Тому, за властивістю 3 $F^*(x)=0$ при $x \le 1$ і $F^*(x)=1$ при $x > 6$.

а) Значення $X < 4$, а саме $x_1 = 1$ спостерігалося 10 разів, тому

$$F^*(x) = \frac{10}{50} = 0{,}2 \text{ при } 1 < x \le 4.$$

б) Значення $X < 6$, а саме $x_1 = 1$ та $x_2 = 4$ спостерігалося 10+15=25 разів, звідси

$$F^*(x) = \frac{25}{50} = 0{,}5 \text{ при } 4 < x \le 6.$$

Отже, шукана емпірична функція розподілу має вигляд

$$F^*(x) = \begin{cases} 0 \;\; при \;\; x \le 1; \\ 0{,}2 \;\; при \;\; 1 < x \le 4; \\ 0{,}5 \;\; при \;\; 4 < x \le 6; \\ 1 \;\; при \;\; x > 6, \end{cases}$$

а її графік зображено на рис. 6.8.

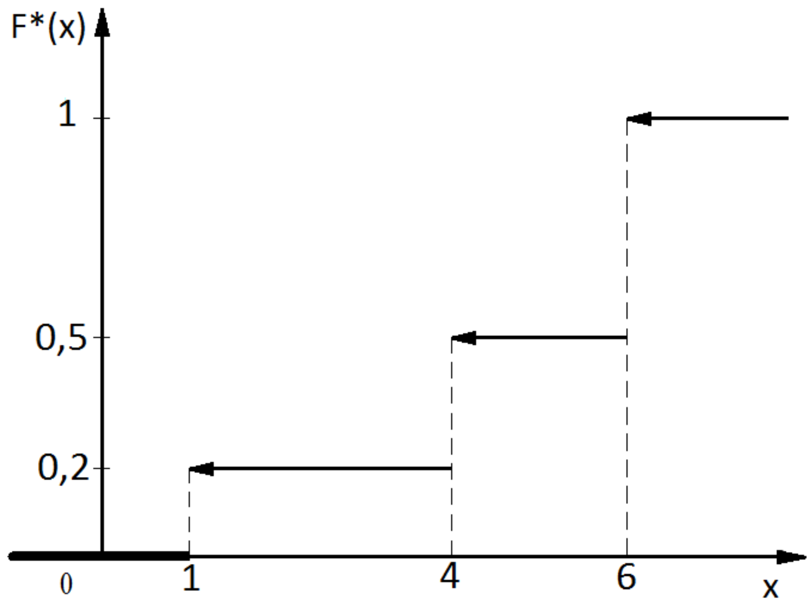

Рис. 6.8.

Емпірична функція є кусково-сталою функцією. Графік $F^*(x)$ являє сходинкову лінію, яка у напрямку числової осі змінює свої значення від нуля до одиниці.

6.4. Статистичні оцінки параметрів та їх властивості.

Нехай потрібно вивчити кількісну ознаку X генеральної сукупності. Припустимо, що з теоретичних міркувань встановлено, який саме розподіл має ознака X. Природно виникає питання оцінювання параметрів, якими визначається цей розподіл. Наприклад, якщо відомо, що розподіл в генеральній сукупності нормальний, то необхідно оцінити (наближено знайти) $M(X)=a$ і $\sigma(X)=\sigma$, якщо є підстави вважати, що X має показниковий розподіл − оцінити параметр λ.

Нехай з генеральної сукупності з функцією розподілу $F(x,\theta)$, де θ − невідомий параметр, здійснено вибірку об'єму n:

$$x_1,\ x_2,\ \ldots,\ x_n. \tag{6.1}$$

Це значення величини X, що одержані в результаті n незалежних спостережень. Для знаходження наближених значень (оцінки) невідомого параметра θ будемо розглядати функції вигляду

$$\theta^* = f\ (x_1, x_2, \ldots, x_n), \tag{6.2}$$

які називаються вибірковими функціями або статистиками.

Задача оцінки невідомого параметра θ зводиться до знаходження таких вибіркових функцій (6.2), які можна використовувати як оцінку параметра θ.

Припустимо, що (6.2) − статистична оцінка невідомого параметра θ теоретичного розподілу та нехай за допомогою вибірки об'єму n (6.1) знайдено оцінку θ_1^*. Повторимо дослід, тобто візьмемо з генеральної сукупності іншу вибірку того ж об'єму n і на її основі знайдемо оцінку θ_2^*. Повторюючи дослід багатократно, дістанемо числа θ_1^*, θ_2^*, …, θ_k^*, що будуть в загальному випадку різними. Оскільки кожна вибірка випадкова, то оцінку θ^* можна розглядати як випадкову величину, а числа θ_1^*, θ_2^*, …, θ_k^* − як її можливі значення.

Для того, щоб оцінки давали «добрі» наближення параметрів, вони мають задовольняти певним вимогам, а саме: бути незміщеними, ефективними та спроможними.

Означення 1. Оцінку θ^* невідомого параметра θ називають незміщеною, якщо її математичне сподівання дорівнює оцінюваному параметру:

$$M(\theta^*) = \theta, \tag{6.3}$$

де $M(\theta^*)$ − математичне сподівання оцінки θ^*.

Якщо умова (6.3) не виконується, оцінка θ^* параметра θ називається зміщеною. Вимога незміщеності гарантує відсутність систематичних (одного знаку) помилок при оцінці параметрів.

Незміщена оцінка θ^* дає хороший результат, якщо її можливі значення мало розсіяні навколо свого середнього значення, тобто дисперсія $D(\theta^*)$ мала.

Означення 2. Статистична оцінка θ^* називається ефективною, якщо вона має найменшу дисперсію серед всіх можливих оцінок параметра θ, обчислених за вибірками одного й того ж об'єму.

При розгляді вибірок великого об'єму статистичні оцінки мають задовольняти умову спроможності.

Означення 3. Статистичну оцінку θ^* параметра θ називають спроможною, якщо при збільшенні кількості незалежних випробувань n вона прямує за ймовірністю до значення параметра θ

$$\lim_{n\to\infty} P(|\theta^* - \theta| < \varepsilon) = 1. \tag{6.4}$$

для довільного $\varepsilon > 0$.

Зокрема, якщо дисперсія незміщеної оцінки при $n \to \infty$ прямує до нуля, то така оцінка виявляється і спроможною.

Розрізняють точкові та інтервальні оцінки параметрів.

Точковою називають оцінку невідомого параметра θ, що визначається одним числом.

Інтервальною називають оцінку, яка визначається двома числами − кінцями інтервалу, відносно якого з наперед заданою ймовірністю можна стверджувати, що оцінюваний параметр знаходиться всередині цього інтервалу.

Зазначимо, що при вибірках малого об'єму точкові оцінки можуть приводити до грубих помилок. Тоді застосовують інтервальні оцінки.

6.5. Точкові оцінки параметрів розподілу.

Нехай вивчається ознака X генеральної сукупності за результатами вибірки x_1, x_2, ..., x_n об'єму n.

Для побудови точкових оцінок математичного сподівання $M(X)=a$ і дисперсії $D(X)=\sigma^2$ випадкової величини X (їх називають також генеральною середньою $\bar{x}_Г$ та генеральною дисперсією $D_Г$) використовують відповідні вибіркові характеристики (середні показники).

6.5.1. Вибіркова середня.

Означення 1. Вибірковою середньою $\bar{x}_B$ називають середнє арифметичне значень вибірки x_1, x_2, …, x_n:

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \qquad (6.5)$$

Якщо значення ознаки X : x_1, x_2, …, x_k мають відповідно частоти $n_1, n_2, ..., n_k$; причому $n_1 + n_2 + ... + n_k = n$, n − об'єм вибірки, то

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \ . \qquad (6.6)$$

В якості оцінки генеральної середньої приймають вибіркову середню. Така оцінка є незміщеною та спроможною.

Ефективність оцінки залежить від закону розподілу ознаки X. Так, доведено, що коли випадкова величина X розподілена за нормальним законом, то оцінка $\bar{x}_B$ буде ефективною.

6.5.2. Вибіркова та "виправлена" дисперсія.

Щоб охарактеризувати розсіювання вибіркових значень навколо свого середнього значення $\bar{x}_B$ вводять показник «вибіркова дисперсія».

Означення 2. Вибірковою дисперсією D_B називається середнє арифметичне квадратів відхилень вибіркових значень X від вибіркової середньої $\bar{x}_B$.

а) Якщо всі варіанти x_1, x_2, …, x_n різні (не згруповані дані), то

$$D_B = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}_B)^2 \ . \qquad (6.7)$$

б) Для згрупованих даних (значення ознаки x_1, x_2, …, x_k мають відповідно частоти $n_1, n_2, ..., n_k$; причому $n_1 + n_2 + ... + n_k = n$) вибіркова дисперсія D_B обчислюється так:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}_B\right)^2 . \qquad (6.8)$$

Обчислення вибіркової дисперсії D_B можна значно спростити, якщо користуватися формулою

$$D_B = \overline{x^2} - [\bar{x}]^2 \qquad (6.9)$$

(дисперсія дорівнює середньому квадратів значень ознаки X мінус квадрат загальної середньої).

Вибіркова дисперсія D_B є спроможною, але зміщеною оцінкою генеральної дисперсії $D_Г$:

$$M[D_B] = \frac{n-1}{n} D_Г . \qquad (6.10)$$

Користуючись оцінкою D_B ми будемо допускати деяку систематичну похибку у бік менших значень: $M[D_B] < D_Г$. За формулою (6.10) легко «підправити» вибіркову дисперсію так, щоб її математичне сподівання дорівнювало $D_Г$. Для цього достатньо помножити D_B на дріб $\frac{n}{n-1}$. У зв'язку з цим вводять так звану виправлену дисперсію S^2:

$$S^2 = \frac{n}{n-1} D_B . \qquad (6.11)$$

Виправлена дисперсія S^2 вже є незміщеною оцінкою генеральної дисперсії $D_Г$: $M[S^2] = D_Г$.

Для оцінки середнього квадратичного відхилення генеральної сукупності використовують „виправлене" середнє квадратичне відхилення

$$S = \sqrt{S^2} . \qquad (6.12)$$

Така оцінка є зміщеною.

Зауваження. При достатньо великому об'ємі вибірки ($n \geq 30$) вибіркова D_B та "виправлена" дисперсія S^2 відрізняються мало (вже при $n = 30$ на 3%) і тому при $n \geq 30$ вибіркова дисперсія D_B може слугувати оцінкою генеральної дисперсії $D_Г$. На практиці виправленою дисперсію S^2 користуються при $n < 30$.

Приклад 6.5. Вибіркова сукупність задана законом розподілу вибірки

X	1	2	3	4
n_i	20	15	10	5

Знайти вибіркову дисперсію D_B та виправлену дисперсію S^2.

Розв’язання. 1 спосіб. Знайдемо об’єм вибірки: $n = 20 + 15 + 10 + 5 = 50$. Оскільки дані згруповані, вибіркову середню $\bar{x}_B$ обчислюємо за формулою (6.6)

$$\bar{x}_B = \frac{20 \cdot 1 + 15 \cdot 2 + 10 \cdot 3 + 5 \cdot 4}{50} = \frac{100}{50} = 2,$$

а вибіркову дисперсію D_B − за (6.8)

$$D_B = \frac{20 \cdot (1-2)^2 + 15 \cdot (2-2)^2 + 10 \cdot (3-2)^2 + 5 \cdot (4-2)^2}{50} = \frac{20 \cdot 1 + 15 \cdot 0 + 10 \cdot 1 + 5 \cdot 4}{50} = \frac{50}{50} = 1.$$

2 спосіб. Для розрахунку вибіркової дисперсії D_B скористаємося формулою (6.9).

Складемо розподіл квадратів значень ознаки X:

x_i^2	1	4	9	16
n_i	20	15	10	5

та обчислимо середню квадратів значень ознаки X:

$$\overline{x^2} = \frac{20 \cdot 1^2 + 15 \cdot 2^2 + 10 \cdot 3^2 + 5 \cdot 4^2}{50} = \frac{20 + 60 + 90 + 80}{50} = \frac{250}{50} = 5.$$

Отже, $D_B = 5 - 2^2 = 1$ (маємо такий самий результат).

Знаходимо виправлену дисперсію : $S^2 = \frac{50}{49} \cdot 1 \approx 1.02$.

Як бачимо, вибіркова дисперсія D_B та "виправлена" дисперсія S^2 відрізняються мало, оскільки тут об’єм вибірки $n = 50 > 30$.

6.6. Інтервальні оцінки. Довірча ймовірність.

Точкові оцінки параметрів розподілу можна прийняти за початкові орієнтовні результати обстеження спостережень. Їх недоліком є те, що невідомо з якою точністю вони дають оцінюваний параметр θ. Якщо для великого числа спостережень точність, як правило, буває достатньою для практичних висновків, то для вибірок невеликого об’єму питання про точність оцінок стає суттєвим.

З цієї причини при невеликих об’ємах вибірки потрібно використовувати інтервальні оцінки. Задача інтервального оцінювання полягає в наступному.

Нехай знайдена за даними вибірки об’єму n статистична оцінка $\theta^*(x_1, x_2, ..., x_n)$ є точковою характеристикою параметра θ. Очевидно, що θ^*

тим точніше визначає параметр θ, чим меншою є абсолютна величина різниці $|\theta-\theta^*|$. Інакше кажучи, якщо $\delta>0$ і

$$|\theta-\theta^*|<\delta, \tag{6.13}$$

то число δ характеризує точність оцінки. Оскільки θ^* випадкова, а θ − невипадкова величина, то категорично стверджувати, що знайдена оцінка задовольняє нерівності (6.13) неможливо. Можна казати лише про ймовірність γ, з якою ця нерівність виконується.

Означення. Довірчою ймовірністю (надійністю) оцінки θ^* невідомого параметра θ називається ймовірність γ, з якою здійснюється нерівність (6.13), $\delta>0$:

$$P\left(|\theta-\theta^*|<\delta\right)=\gamma. \tag{6.14}$$

Звичайно надійність оцінки γ задається наперед, причому в якості γ беруть число, близьке до одиниці: 0,95; 099; 0,999.

Перетворимо (6.13) за властивостями модуля:

$$|\theta-\theta^*|<\delta,\ -\delta<\theta-\theta^*<\delta$$

або, остаточно,

$$\theta^*-\delta<\theta<\theta^*+\delta.$$

Тому довірчу ймовірність (6.14) можна подати у вигляді:

$$P\left(\theta^*-\delta<\theta<\theta^*+\delta\right)=\gamma. \tag{6.15}$$

Формулу (6.15) слід розуміти так: ймовірність того, що інтервал $\left(\theta^*-\delta,\theta^*+\delta\right)$ містить в собі (покриває) невідомий параметр θ, дорівнює γ. Інтервал $\left(\theta^*-\delta,\theta^*+\delta\right)$ називають довірчим інтервалом. Кінці довірчого інтервалу називають довірчими межами.

Якщо випадкова величина X розподілена нормально із заданим середнім квадратичним відхиленням σ та невідомим математичним сподіванням a, формула (6.15) набуває вигляду

$$P\left(\bar{x}_B-\frac{t\sigma}{\sqrt{n}}<a<\bar{x}_B+\frac{t\sigma}{\sqrt{n}}\right)=\gamma. \tag{6.16}$$

Тут $\delta=\dfrac{t\sigma}{\sqrt{n}}$ − точність оцінки, n − об’єм вибірки, t − значення аргументу функції Лапласа $\Phi(t)$ (див. додатки), при якому $\Phi(t)=\dfrac{\gamma}{2}$.

Якщо ж σ невідоме, то довірчий інтервал визначатиметься так:

$$\left(\bar{x}_B - \frac{t_\gamma S}{\sqrt{n}}; \bar{x}_B + \frac{t_\gamma S}{\sqrt{n}} \right), \qquad (6.17)$$

де S – „виправлене" середнє квадратичне відхилення, а величина $t_\gamma = t(\gamma, n)$ знаходиться за відомими n, γ за допомогою таблиці значень розподілу Стьюдента (див. додатки).

Приклад 6.6. За даним розподілом вибірки

межі інтер валів	[12-14[	[14-16[	[16-18[	[18-20[	[20-22[
n_i	2	7	3	7	1

оцінити з надійністю 0,95 математичне сподівання a нормально розподіленої ознаки X генеральної сукупності за допомогою довірчого інтервалу.

Розв'язання. Інтервальною оцінкою з надійністю $\gamma = 0{,}95$ математичного сподівання a нормально розподіленої ознаки X при невідомому середньому квадратичному відхиленні σ генеральної сукупності (та об'ємі вибірки $n < 30$) слугує довірчий інтервал (6.17).

Знайдемо вибіркову середню $\bar{x}_B$ та «виправлене" середнє квадратичне відхилення S відповідно за формулами

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} n_i x_i \ ,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} n_i (x_i - \bar{x}_B)^2 .$$

В якості значень x_i приймемо середини відповідних часткових інтервалів:

$$\bar{x}_B = \frac{1}{20}(2\cdot 13 + 7\cdot 15 + 3\cdot 17 + 7\cdot 19 + 1\cdot 21) = \frac{336}{20} = 16{,}8;$$

$$S^2 = \frac{1}{19}(2(13-16{,}8)^2 + 7(15-16{,}8)^2 + 3(17-16{,}8)^2 + 7(19-16{,}8)^2 + 1(21-16{,}8)^2) =$$

$$= \frac{1}{19}[2\cdot 14{,}44 + 7\cdot 3{,}24 + 3\cdot 0{,}04 + 7\cdot 4{,}84 + 17{,}64] ==$$

$$= \frac{1}{19}[28{,}88 + 22{,}68 + 0{,}12 + 33{,}88 + 17{,}64] = \frac{103{,}2}{19} \approx 5{,}43.$$

Звідси $S = \sqrt{S^2} = \sqrt{5{,}43} \approx 2{,}33$. Параметр t_γ знаходимо за готовими таблицями (див. додатки) за даними n та γ $(n = 20, \gamma = 0{,}95)$.

Підставивши $\bar{x}_B = 16{,}8;\ \ t_\gamma = 2{,}093;\ \ S = 2{,}33;\ \ n = 20$ у вираз (6.17), дістанемо шуканий довірчий інтервал 15,71< a <17,89, що покриває невідоме математичне сподівання a з надійністю 0,95.

Приклад 6.7. За даними прикладу 6.6 оцінити невідоме математичне сподівання a з надійністю 0,95, якщо середнє квадратичне відхилення $\sigma = 2$.

Розв'язання. На відміну від попередньої, в цій задачі σ – відоме. Тому довірчий інтервал знаходиться вже за формулою:

$$\bar{x}_B - t\frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + t\frac{\sigma}{\sqrt{n}}, \tag{6.18}$$

де $\bar{x}_B = 16{,}8;\quad n = 20;\quad \sigma = 2.$

Отже, усі величини, крім t, відомі. Знайдемо t за співвідношенням $\Phi(t) = \dfrac{\gamma}{2}$, де $\Phi(t)$ − функція Лапласа (див. додатки). Дістанемо: $\Phi(t) = \dfrac{0{,}95}{2} = 0{,}475$, а значення аргументу функції $\Phi(t)$ знаходимо за додатком 2: $t = 1{,}96.$

Підставивши усі дані в формулу (6.18), дістанемо шуканий довірчий інтервал: 15,9< a <17,7.

Практичне заняття №7. Основи математичної статистики

Опитування з теорії.

1. Як означається генеральна сукупність? вибірка?
2. У чому полягає суть вибіркового методу?
3. Що називають варіантами? варіаційним рядом? розмахом вибірки?
4. Що називають частотою, відносною частотою варіант?
5. Як означається статистичний розподіл вибірки? Як побудувати дискретний варіаційний ряд? інтервальний варіаційний ряд?
6. Що являють собою полігон (відносних) частот, гістограма (відносних) частот вибірки?
7. Як означається емпірична функція розподілу? Навести основні властивості.
8. Що розуміють під статистичною оцінкою параметра розподілу?
9. Яка статистична оцінка називається незміщеною? ефективною? спроможною?
10. Яка статистична оцінка називається точковою? Інтервальною?
11. Навести формули для обчислення точкових оцінок математичного сподівання, дисперсії, середнього квадратичного відхилення
12. Як «підправити» вибіркову дисперсію, щоб оцінка стала незміщеною?
13. При яких об’ємах вибірки в якості оцінки генеральної дисперсії приймають "виправлену" дисперсію? Вибіркову дисперсію?
14. Як означається довірча ймовірність (надійність) статистичної оцінки?
15. Як знайти довірчий інтервал для математичного сподівання нормального розподілу при відомому середньому квадратичному відхиленні σ? При невідомому середньому квадратичному відхиленні σ?
16. Як можна підвищити точність статистичної оцінки?
17. За якою формулою знаходиться мінімальний об’єм вибірки для оцінки математичного сподівання з наперед заданою точністю δ і надійністю γ?

Завдання для аудиторної роботи.

Приклад 6.1. Вибірка об’єму $n = 20$ задана у вигляді розподілу відносних частот

X	2	6	12
ω_i	0,15	0,5	0,35

Скласти статистичний ряд розподілу частот.

Розв’язання. Знайдемо частоти, що відповідають відносним частотам

$$n_1 = \omega_1 \cdot n = 0{,}15 \cdot 20 = 3;\;\; n_2 = \omega_2 \cdot n = 0{,}5 \cdot 20 = 10;\;\; n_3 = \omega_3 \cdot n = 0{,}35 \cdot 20 = 7\,.$$

Запишемо статистичний ряд розподілу частот:

X	2	6	12
n_i	3	10	7

Контроль: $3+10+7=20=n$ (сума всіх частот дорівнює об'єму вибірки).

Приклад 6.2. У цеху встановлено 7 верстатів. Протягом 25 днів реєструвалась кількість верстатів, які не працювали. Одержано такі значення: 0,1,2,1,1,7,3,2,1,4,2,0,0,7,2,3,3,1,0,1,2,1,3,5,0. Записати вибірку у вигляді: а) варіаційного ряду; б) статистичного ряду частот і відносних частот. Знайти розмах вибірки.

Розв'язання. а) Запишемо значення варіант у зростаючому порядку. Дістанемо варіаційний ряд вигляду:

0,0,0,0,0,1,1,1,1,1,1,1,2,2,2,2,2,3,3,3,3,4,5,7,7;

б) На підставі вибіркових даних складемо статистичний ряд частот і відносних частот (об'єм вибірки $n=25$):

X	0	1	2	3	4	5	7
n_i	5	7	5	4	1	1	2
$\omega_i=\frac{n_i}{n}$	$\frac{5}{25}=0,2$	$\frac{7}{25}=0,28$	$\frac{5}{25}=0,2$	$\frac{4}{25}=0,16$	$\frac{1}{25}=0,04$	$\frac{1}{25}=0,04$	$\frac{2}{25}=0,08$

Контроль: $0,2+0,28+0,2+0,16+0,04+0,04+0,08=1$.

Розмах вибірки $x_{\max}-x_{\min}=7-0=7$.

Приклад 6.3. За статистичним рядом розподілу прикладу 6.2 побудувати інтервальний ряд, розбивши діапазон значень на інтервали довжиною $h=1,5$. Знайти щільності відносних частот.

Розв'язання. Занесемо дані в таблицю

Інтервал	[0; 1,5)	[1,5; 3)	[3; 4,5)	[4,5; 6)	[6; 7,5)
n_i	5+7=12	5+4=9	1	1	2
$\omega_i=\frac{n_i}{n}$	$\frac{12}{25}=0,48$	$\frac{9}{25}=0,36$	$\frac{1}{25}=0,04$	$\frac{1}{25}=0,04$	$\frac{2}{25}=0,08$
$\frac{\omega_i}{h}$	0,32	0,24	$\approx$0,027	$\approx$0,027	$\approx$0,053

Контроль: $0,48+0,36+0,04+0,04+0,08=1$ (сума відносних частот дорівнює одиниці).

Приклад 6.4. Для групованої вибірки з довжиною інтервалу $h=30$

Інтервал	[50; 80)	[80; 110)	[110; 140)	[140; 170)
ω_i	0,1	0,3	0,5	0,1

побудувати: а) нормовану стовпчикову діаграму; б) гістограму відносних частот.

Розв'язання. За формулою $\frac{\omega_i}{h}$ $i = 1,2,3,4$ обчислимо щільності відносних частот та доповнимо вихідну таблицю:

Інтервал	[50; 80)	[80; 110)	[110; 140)	[140; 170)
ω_i	0,1	0,3	0,5	0,1
$\frac{\omega_i}{h}$	$\frac{0,1}{30}$	$\frac{0,3}{30}$	$\frac{0,5}{30}$	$\frac{0,1}{30}$

Далі по осі OX відкладають частинні інтервали, а над ними проводять відрізки, паралельні осі абсцис на відстані $\frac{\omega_i}{h}$, $i = 1,2,3,4$ відповідно (рис. 6.1).

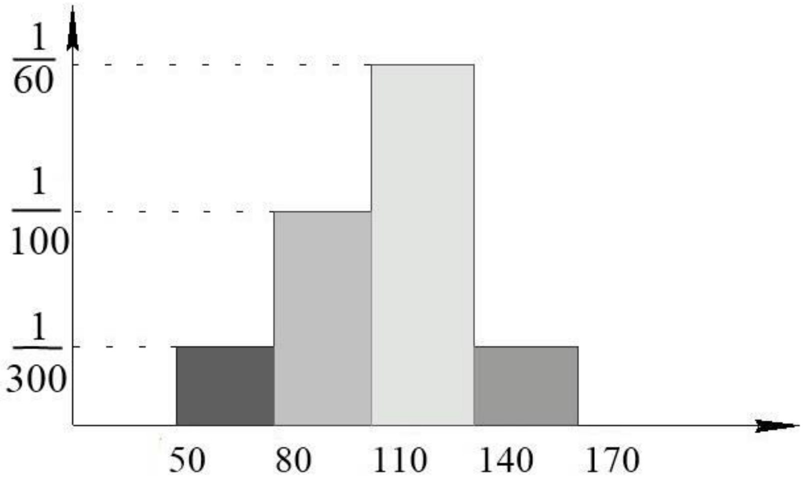

Рис. 6.1.

За нормованою стовпчиковою діаграмою будують гістограму відносних частот (рис. 6.2).

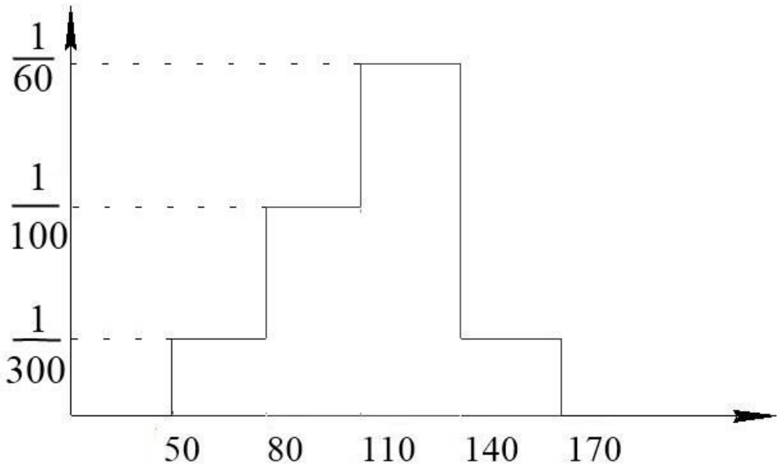

Рис. 6.2.

Приклад 6.5. Для ряду розподілу,

Ф (флегматик)	М (меланхолік)	Х (холерик)	С (сангвінік)
0,48	0,12	0,2	0,2

що відповідає розподілу за темпераментом (дискретні дані) побудувати секторну діаграму.

Розв’язання. Для побудови секторної діаграми зробимо додаткові розрахунки:

$\alpha_{\Phi} = 360^{\circ} \cdot 0{,}48 = 172{,}8^{\circ}$; $\alpha_M = 360^{\circ} \cdot 0{,}12 = 43{,}2^{\circ}$; $\alpha_X = 360^{\circ} \cdot 0{,}2 = 72^{\circ}$; $\alpha_C = 360^{\circ} \cdot 0{,}2 = 72^{\circ}$.

За цими даними одержимо секторну діаграму такого вигляду:

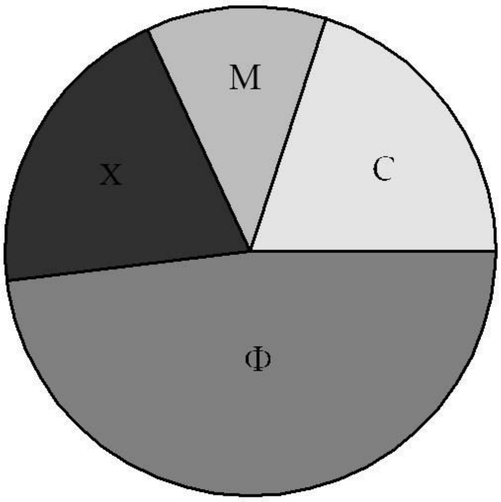

Рис. 6.3.

Приклад 6.6. За рядом розподілу відносних частот

X	-1	1	2	3
$\omega_i = \frac{n_i}{n}$	0,2	0,3	0,1	0,4

обчислити вибіркову середню і дисперсію.

Розв’язання. Оскільки дані згруповані, вибіркову середню $\bar{x}_B$ обчислюємо за формулою (6.6) або $\bar{x}_B = \sum_{i=1}^{k} x_i \frac{n_i}{n} = \sum_{i=1}^{k} x_i \omega_i$:

$$\bar{x}_B = -1 \cdot 0{,}2 + 1 \cdot 0{,}3 + 2 \cdot 0{,}1 + 3 \cdot 0{,}4 = 1{,}5\,.$$

Так само перетворимо формулою (6.8) для розрахунку дисперсії D_B:

$$D_B = \sum_{i=1}^{k} \frac{n_i}{n} (x_i - \bar{x}_B)^2 = \sum_{i=1}^{k} \omega_i (x_i - \bar{x}_B)^2\,.$$

Отже, маємо

$$D_B = (-1-1{,}5)^2 \cdot 0{,}2 + (1-1{,}5)^2 \cdot 0{,}3 + (2-1{,}5)^2 \cdot 0{,}1 + (3-1{,}5)^2 \cdot 0{,}4 = 2{,}25 .$$

Приклад 6.7. За вибіркою об'єму $n = 7$ знайдено зміщену оцінку генеральної дисперсії $D_в = 4$. Знайти незміщену оцінку генеральної дисперсії.

Розв'язання. За формулою (6.11) знайдемо виправлену дисперсію S^2

$$S^2 = \frac{n}{n-1} D_B = \frac{7}{6} \cdot 4 \approx 4{,}67 ,$$

що і слугує незміщеною оцінкою генеральної дисперсії.

Приклад 6.8. Проведено 5 рівно точних вимірювань одним приладом відстані від гармати до цілі, $\sigma = 40$ м (середнє квадратичне відхилення). Знайти довірчий інтервал для оцінки «істинної» відстані a до цілі з надійністю 0,95, якщо середнє арифметичне результатів вимірювань $\bar{x}_B = 2000$ м (припускається, що результати розподілені нормально).

Розв'язання. Довірчий інтервал для оцінки математичного сподівання a величини X − відстані від гармати до цілі знайдемо за формулою (6.18)

$$\bar{x}_B - t\frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + t\frac{\sigma}{\sqrt{n}} ,$$

оскільки σ відоме.

Для нашої задачі $\bar{x}_B = 2000; \quad n = 5; \quad \sigma = 40.$

Отже, усі величини, крім t, відомі. Знайдемо t за співвідношенням $\Phi(t) = \frac{\gamma}{2}$, де $\Phi(t)$ − функція Лапласа (див. додаток 2). Дістанемо: $\Phi(t) = \frac{0{,}95}{2} = 0{,}475,$ а значення аргументу функції $\Phi(t)$ знаходимо за додатком 2: $t = 1{,}96.$

Тепер обчислимо $\delta = \frac{t\sigma}{\sqrt{n}}$ − точність оцінки: $\delta = \frac{1{,}96 \cdot 40}{\sqrt{5}} \approx 35{,}06$.

Підставивши усі дані в формулу (6.18), дістанемо шуканий довірчий інтервал: 1964,94< a <2035,06.

Шуканий інтервал з надійністю $\gamma = 0{,}95$ покриває невідомий параметр $M(X) = a$ (тобто у 95% параметр знаходиться всередині цього інтервалу).

Приклад 6.9. За вибіркою об'єму $n = 10$:

X	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

оцінити з надійністю 0,95 математичне сподівання a нормально розподіленої ознаки X генеральної сукупності за вибірковою середньою, побудувавши довірчий інтервал.

Розв'язання. Вибіркову середню та „виправлене” середнє квадратичне відхилення знайдемо відповідно за формулами:

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i n_i\ ,\ S^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i\left(x_i - \bar{x}_B\right)^2\ ,\ S = \sqrt{S^2}\ .$$

Підставивши у ці формули дані задачі, дістанемо

$$\bar{x}_B = \frac{-2\cdot 2 + 1\cdot 1 + 2\cdot 2 + 3\cdot 2 + 4\cdot 2 + 5\cdot 1}{10} = \frac{20}{10} = 2\,,$$

$$S^2 = \frac{2\cdot(-2-2)^2 + 1\cdot(1-2)^2 + 2\cdot(2-2)^2 + 2\cdot(3-2)^2 + 2\cdot(4-2)^2 + 1\cdot(5-2)^2}{9} =$$

$$= \frac{2\cdot 16 + 1\cdot 1 + 2\cdot 1 + 2\cdot 4 + 1\cdot 9}{9} = \frac{52}{9} \approx 5{,}78\,,$$

$$S = \sqrt{5{,}78} \approx 2{,}4\,.$$

За таблицею розподілу Стьюдента (додаток 3) знайдемо величину t_γ за відомими $\gamma = 0{,}95$ і $n = 10$: $t_\gamma = 2{,}26$. Довірчий інтервал для математичного сподівання a знайдемо за формулою (6.17):

$$\left(\bar{x}_B - \frac{t_\gamma S}{\sqrt{n}}; \bar{x}_B + \frac{t_\gamma S}{\sqrt{n}}\right).$$

Отже, маємо $\left(2 - \frac{2{,}26\cdot 2{,}4}{\sqrt{10}};\ 2 + \frac{2{,}26\cdot 2{,}4}{\sqrt{10}}\right)$ або $\left(0{,}3;\ 3{,}7\right)$. Шуканий довірчий інтервал покриває математичне сподівання a з надійністю $\gamma = 0{,}95$.

Завдання для самостійної роботи.

1. Вибірка задана у вигляді ряду розподілу частот:

x_i	1	4	5	7
n_i	20	10	14	6

Знайти розподіл відносних частот. Побудувати полігон відносних частот.

Відповідь:

x_i	1	4	5	7
$\omega_i = \frac{n_i}{n}$	2/5	1/5	7/25	3/25

2. Вибірка об'єму $n = 50$ задана у вигляді розподілу відносних частот

x_i	-1	0	2	4	5
$\omega_i = \frac{n_i}{n}$	0,2	0,3	0,2	0,1	0,2

Записати статистичний ряд розподілу частот.
Відповідь:

x_i	-1	0	2	4	5
n_i	10	15	10	5	10

3. Вибірка задана у вигляді ряду розподілу частот

x_i	-3	-1	2	3
n_i	10	15	20	5

Знайти емпіричну функцію розподілу.

Відповідь: $F^*(x) = \begin{cases} 0, x < -3 \\ 0{,}2, -3 \le x < -1 \\ 0{,}5, -1 \le x < 2 \\ 0{,}9, 2 \le x < 3 \\ 1, x \ge 3 \end{cases}$.

4. Побудувати полігон частот та полігон відносних частот за даним розподілом вибірки

x_i	1	4	5	7
n_i	20	10	14	6

5. Побудувати гістограму відносних частот за даним розподілом вибірки

Інтервал	[1; 5)	[5; 9)	[9; 13)	[13;17)	[17;21)
n_i	10	20	50	12	8

6. За статистичним розподілом вибірки

x_i	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

знайти незміщені оцінки для генеральної середньої та генеральної дисперсії.
Відповідь: $\bar{x}_B$=4,004; S^2=0,01606.

7. Сумарне число балів, набраних у змаганнях, має вигляд інтервального ряду:

Інтер вал	4,5—5,5	5,5—6,5	6,5—7,5	7,5—8,5	8,5—9,5	9,5—10,5	10,5—11,5	11,5—12,5	12,5—13,5	13,5—14,5
n_i	40	32	28	24	20	18	16	12	8	4

Знайти основні вибіркові характеристики (середню, дисперсію, середнє квадратичне відхилення.
Відповідь: $\bar{x}_B$=8,02; $D_B \approx$6,37; $S^2 \approx$6,4; $S = \sqrt{6{,}4} \approx 2{,}53$.

8. П'ять вимірювань довжини стержня (в мм) одним приладом (без систематичних помилок) дали такі результати: 92, 94, 103, 105, 106.

Знайти: а) вибіркову середню довжини стержня; б) вибіркову і виправлену дисперсії похибок приладу.
Відповідь: а) $\bar{x}_B$=100; б) D_B=34; S^2=42,5.

9. За даними десяти вимірювань деякої фізичної величини знайдені середнє арифметичне результатів вимірювань $\bar{x} = 3{,}4$ та виправлене середнє квадратичне відхилення $s = 2{,}4$. Оцінити істинне значення фізичної величини з надійністю $\gamma = 0{,}95$.
Відповідь: $]1{,}7; 5{,}1[$.

10. З нормально розподіленої генеральної сукупності зроблена вибірка об'єму $n = 15$:

x_i	-3	-2,5	-2	-1,5	-0,5	1
n_i	1	2	4	3	3	2

Оцінити математичне сподівання з надійністю $\gamma = 0{,}95$ за допомогою довірчого інтервалу.
Відповідь: $]-2; -0{,}67[$.

11. Випадкова величина розподілена за нормальним законом. На підставі дванадцяти вимірювань знайдені середнє арифметичне результатів вимірювань: $\bar{x} = 1{,}2$ та вибіркова дисперсія: $D_в = 1{,}4$. Знайти довірчий інтервал для математичного сподівання з надійністю $\gamma = 0{,}99$.
Відповідь: $]0{,}1; 2{,}3[$.

12. Час X безвідмовної роботи електронної лампи − випадкова величина, розподілена за нормальним законом з параметром $\sigma = \sigma(X) = 10$ год. Знайти довірчий інтервал для оцінки з надійністю $\gamma = 0{,}99$ середнього часу безвідмовної роботи лампи, якщо вибіркова середня, обчислена за вибіркою об'єму $n = 100$, дорівнює $\bar{x}_B$=500.
Відповідь: $]497{,}42;\ 502{,}58[$.

13. Ємність конденсатора X − випадкова величина, розподілена за нормальним законом, причому $\sigma=\sigma(X)=4$мкФ. Знайти довірчий інтервал для оцінки з надійністю $\gamma=0{,}99$ середньої ємності конденсатора, якщо вибіркова середня, обчислена за вибіркою об'єму $n=16$, дорівнює $\bar{x}_B$=20мкФ.
Відповідь: $]17{,}42;\ 22{,}58[$.

14. Знайти довірчий інтервал для оцінки з надійністю γ невідомого математичного сподівання a нормально розподіленої ознаки X генеральної сукупності, якщо задані генеральне середнє квадратичне відхилення σ, вибіркова середня $\bar{x}_B$ та об'єм вибірки n:

а) $\sigma=4$; $\bar{x}_B=10{,}2$; $n=16$; $\gamma=0{,}999$;

б) $\sigma=5$; $\bar{x}_B=16{,}8$; $n=25$; $\gamma=0{,}99$.

Відповідь: а) $]6{,}89;\ 13{,}51[$; б) $]14{,}23;\ 19{,}37[$.

15. Знайти мінімальний об'єм вибірки, при якому з надійністю γ точність оцінки математичного сподівання a нормально розподіленої ознаки генеральної сукупності буде дорівнювати δ, якщо середнє квадратичне відхилення σ відоме:

а) $\sigma=1{,}2$; $\delta=0{,}3$; $\gamma=0{,}975$;

б) $\sigma=1{,}5$; $\delta=0{,}2$; $\gamma=0{,}925$.

Відповідь: а) 81; б) 179.

Список джерел.

1. Сулима І.М., Яковенко В.М. Вища математика. Теорія ймовірностей. Математична статистика. Навчальний посібник. К.: Вид. Центр НАУ, 2004. – 238 с.
2. Сулима І.М., Ковтун І.І., Нікітіна І.А., Скороход Т.А., Яковенко В.М. Прикладна математика. Теорія ймовірностей. Математична статистика. Навчально-методичний посібник. К.: Вид. Центр НАУ, 2005. – 148 с.
3. Панталієнко Л.А. Методичні вказівки до виконання тестових завдань з дисципліни «Вища математика« за модулем «Теорія ймовірностей, математична статистика та основи кореляційного аналізу». – Видавничий центр НУБІП, 2011. – 71 с.
4. Сулима І.М., Панталієнко Л.А., Яковенко В.М. Методичні рекомендації та індивідуальні завдання з дисципліни „Прикладна математика" для студентів інженерних факультетів. – К.: Вид. центр НАУ, 2001. – 67 с.
5. Медведєв М.Г., Пащенко І.О. Теорія ймовірностей і математична статистика: К.: Ліра-К, 2008. – 536 с.
6. Астахов В.М. Теорія ймовірностей і математична статистика. Навчально-методичний посібник / В.М.Астахов, Г.С. Буланов, В.О. Паламарчук – Краматорськ: ДДМА, 2009. – 64 с.

http://www.dgma.donetsk.ua/metod/vm/tims.pdf

Викладач: доц. Панталієнко Л.А.

Термін виконання: до 1 травня 2020р.

Прошу надсилати виконані завдання за адресою: wnyrk15@gmail.com